Data Section

I have researched and found a data set which looks at the volume and severity of road traffic accidents in Seattle City between 2004 and 2020. The data set is comprised of over 194,000 traffic accidents in Seattle City in that time period and for each accident we have whether the accident was property damage only or whether it also included an injury to a person. There is lots of other information supplied also, this includes the features such as the date of the traffic accident, the location of the accident, the collision type, the number of people involved, the vehicle type, the weather at the time of the accident, if the accident took place in daylight and more. All of these features will be tested to understand if they help to predict the severity of an accident.

Using this information I plan to test a series of machine learning models in order to predict traffic accident severity in Chicago city. I will use a number of machine leaning classification techniques include logistic regression, decision trees and support vector machines in order to understand which model predicts accident severity with the most accuracy. Once the models are built, we will be able to understand which of the features above are most important to predict the severity of road traffic accidents in Seattle City. This will help the health service better plan their resources, for example if they know that injuries are more likely to happen in wet weather they can have more support staff and vehicles ready in wetter months.

Seattle City data source:

Seattle city accident data: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

A descriptive summary of the data source can be found here: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf